1. Project Overview

This project analyzes customer shopping behavior using transactional data from 3,900 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

2. Dataset Summary

- Rows: **3,900**

- Columns: **18**

- Key Features: -

- Customer demographics (Age, Gender, Location, Subscription Status)

- Purchase details (Item Purchased, Category, Purchase Amount, Season, Size, Color)

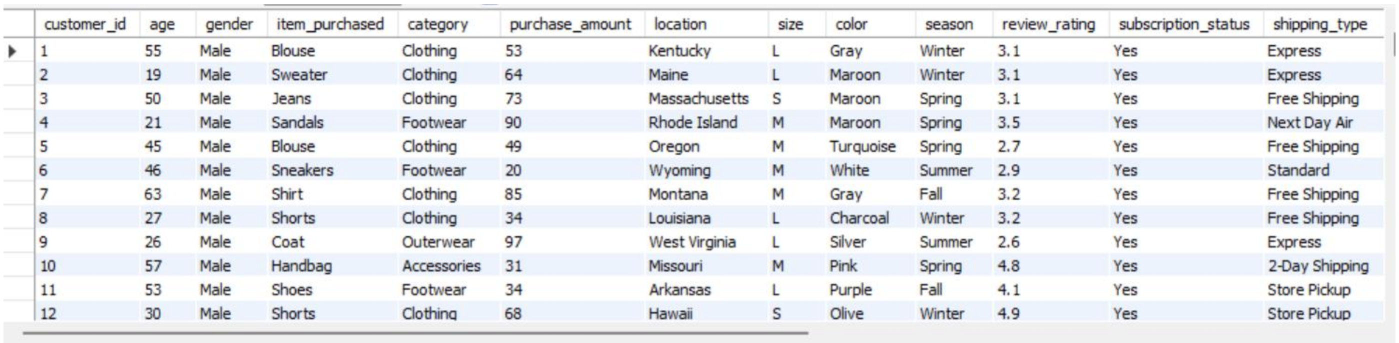
- Shopping behavior (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)

- Missing Data: **37** values in Review Rating column .

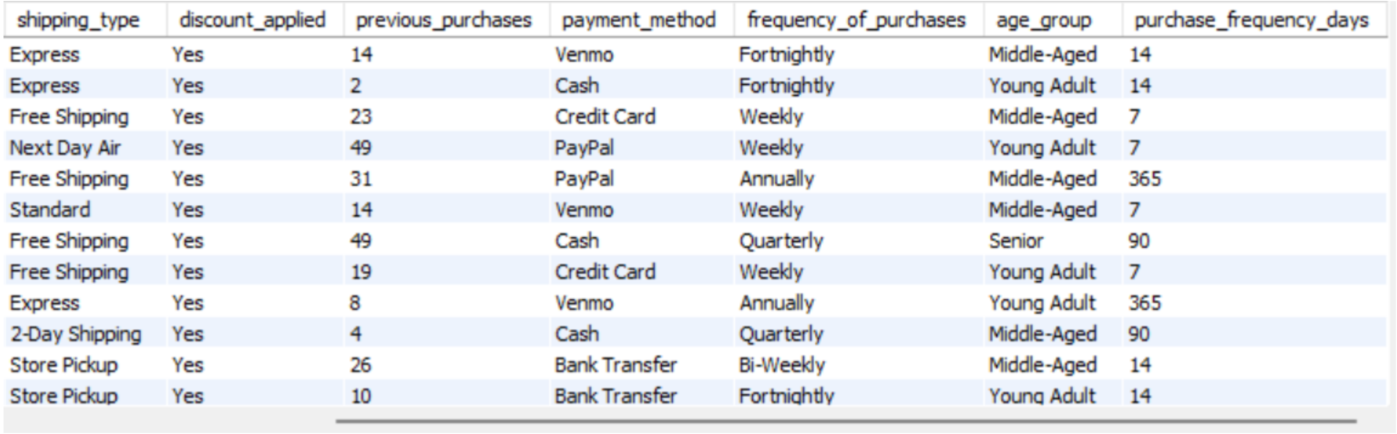
3. Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

- **Data Loading:** Imported the dataset using pandas.
- **Initial Exploration:** Used df.info() to check structure and .describe() for summary statistics.



customer_id	age	gender	item_purchased	category	purchase_amount	location	size	color	season	review_rating	subscription_status	shipping_type
1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express
2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express
3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping
4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air
5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping
6	46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	2.9	Yes	Standard
7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2	Yes	Free Shipping
8	27	Male	Shorts	Clothing	34	Louisiana	L	Charcoal	Winter	3.2	Yes	Free Shipping
9	26	Male	Coat	Outerwear	97	West Virginia	L	Silver	Summer	2.6	Yes	Express
10	57	Male	Handbag	Accessories	31	Missouri	M	Pink	Spring	4.8	Yes	2-Day Shipping
11	53	Male	Shoes	Footwear	34	Arkansas	L	Purple	Fall	4.1	Yes	Store Pickup
12	30	Male	Shorts	Clothing	68	Hawaii	S	Olive	Winter	4.9	Yes	Store Pickup



shipping_type	discount_applied	previous_purchases	payment_method	frequency_of_purchases	age_group	purchase_frequency_days
Express	Yes	14	Venmo	Fortnightly	Middle-Aged	14
Express	Yes	2	Cash	Fortnightly	Young Adult	14
Free Shipping	Yes	23	Credit Card	Weekly	Middle-Aged	7
Next Day Air	Yes	49	PayPal	Weekly	Young Adult	7
Free Shipping	Yes	31	PayPal	Annually	Middle-Aged	365
Standard	Yes	14	Venmo	Weekly	Middle-Aged	7
Free Shipping	Yes	49	Cash	Quarterly	Senior	90
Free Shipping	Yes	19	Credit Card	Weekly	Young Adult	7
Express	Yes	8	Venmo	Annually	Young Adult	365
2-Day Shipping	Yes	4	Cash	Quarterly	Middle-Aged	90
Store Pickup	Yes	26	Bank Transfer	Bi-Weekly	Middle-Aged	14
Store Pickup	Yes	10	Bank Transfer	Fortnightly	Young Adult	14

- **Missing Data Handling:** Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.

- **Column Standardization:** Renamed columns to **snake case** for better readability and documentation.

- **Feature Engineering:**

  ○ Created **age_group** column by binning customer's ages.

  ○ Created **purchase_frequency_days** column from purchase data.

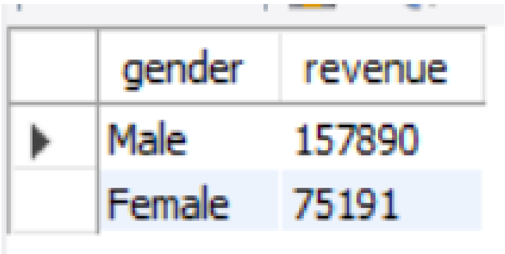
- **Data Consistency Check:** Verified if discount_applied and promo_code_used were redundant; dropped promo_code_used.

- **Database Integration:** Connected Python script to MySQL and loaded the cleaned DataFrame into the database for SQL analysis.

4. Data Analysis using SQL (Business Transactions)

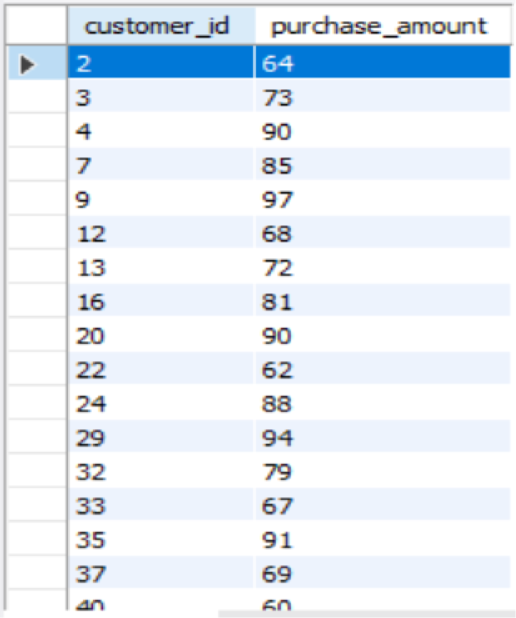
We performed structured analysis in MySQL to answer key business questions:

1. Revenue by Gender – Compared total revenue generated by male vs. female customers.



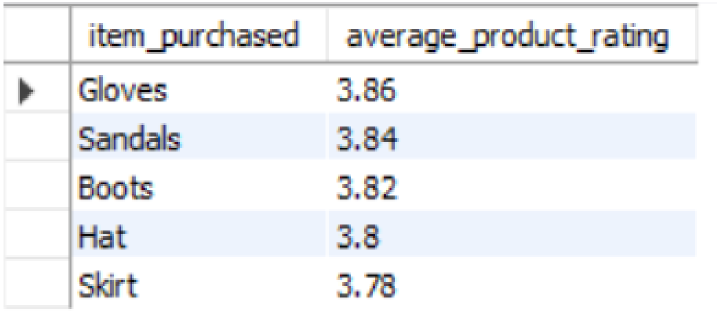
gender	revenue
Male	157890
Female	75191

2. High-Spending Discount Users – Identified customers who used discounts but still spent above the average purchase amount.



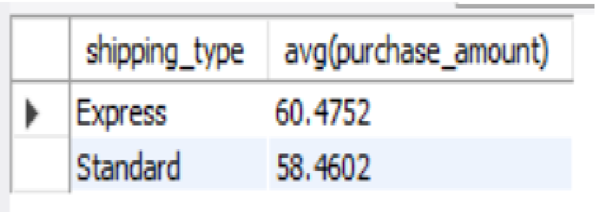
customer_id	purchase_amount
2	64
3	73
4	90
7	85
9	97
12	68
13	72
16	81
20	90
22	62
24	88
29	94
32	79
33	67
35	91
37	69
40	60

3. Top 5 Products by Rating – Found products with the highest average review ratings.



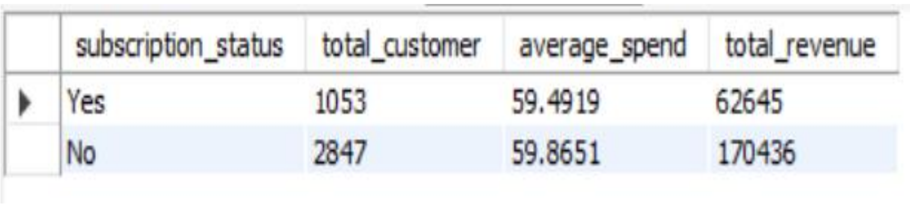
item_purchased	average_product_rating
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.8
Skirt	3.78

4. Shipping Type Comparison – Compared average purchase amounts between Standard and Express shipping.



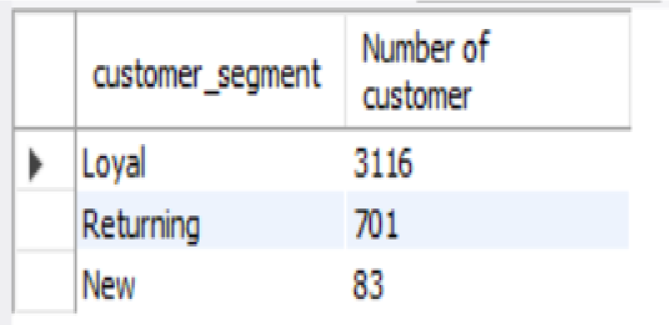
shipping_type	avg(purchase_amount)
Express	60.4752
Standard	58.4602

5. Subscribers vs. Non-Subscribers – Compared average spend and total revenue across subscription status.



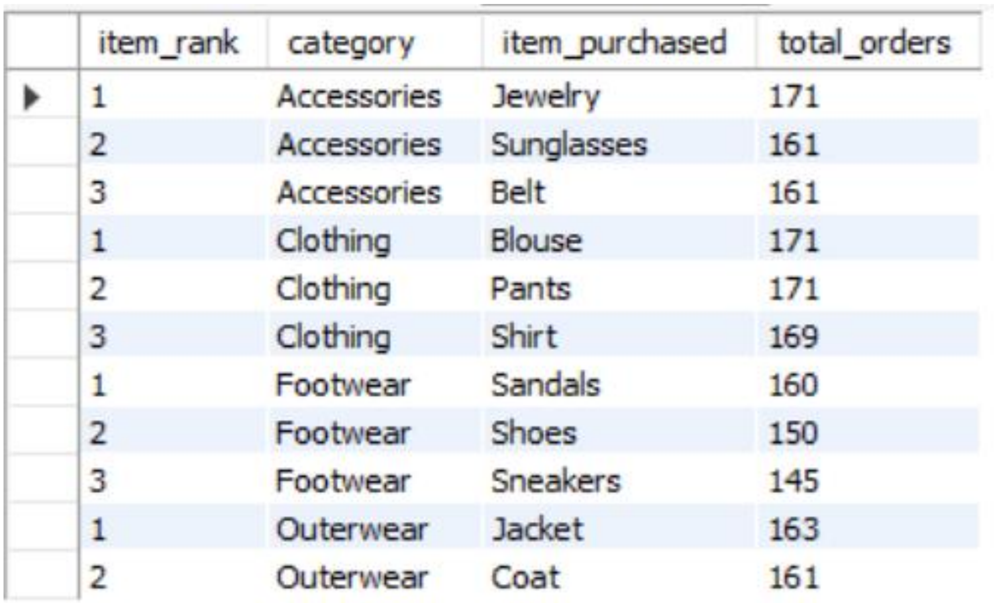
subscription_status	total_customer	average_spend	total_revenue
Yes	1053	59.4919	62645
No	2847	59.8651	170436

6. Customer Segmentation – Classified customers into New, Returning, and Loyal segments based on purchase history.



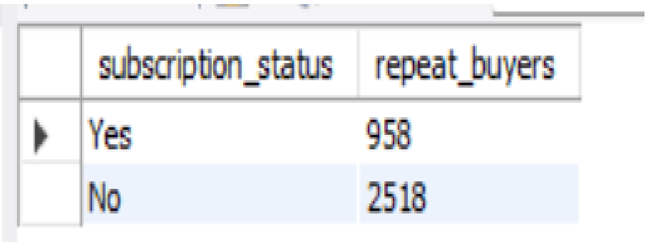
	customer_segment	Number of customer
▶	Loyal	3116
	Returning	701
	New	83

7. Top 3 Products per Category – Listed the top 3 purchased products within each category.



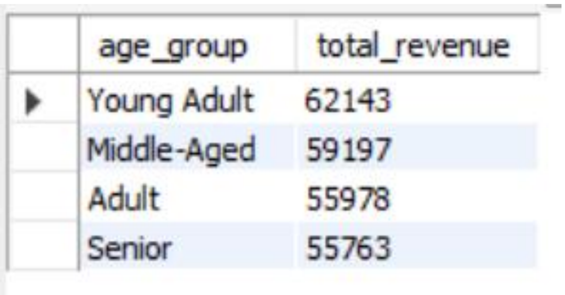
	item_rank	category	item_purchased	total_orders
▶	1	Accessories	Jewelry	171
	2	Accessories	Sunglasses	161
	3	Accessories	Belt	161
	1	Clothing	Blouse	171
	2	Clothing	Pants	171
	3	Clothing	Shirt	169
	1	Footwear	Sandals	160
	2	Footwear	Shoes	150
	3	Footwear	Sneakers	145
	1	Outerwear	Jacket	163
	2	Outerwear	Coat	161

8. Repeat Buyers & Subscriptions – Checked whether customers with >5 purchases are more likely to subscribe.



	subscription_status	repeat_buyers
▶	Yes	958
	No	2518

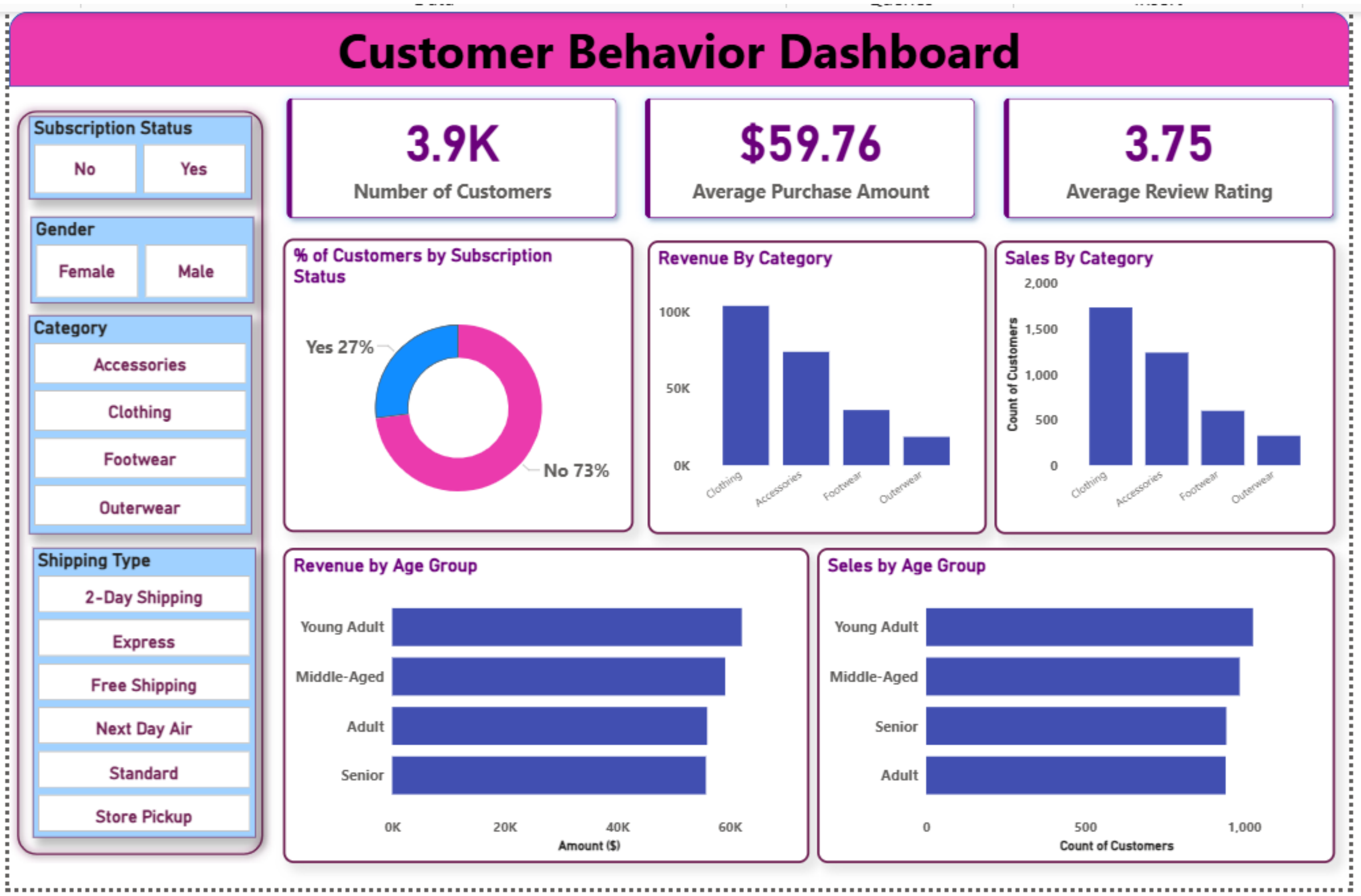
10. Revenue by Age Group – Calculated total revenue contribution of each age group.



age_group	total_revenue
Young Adult	62143
Middle-Aged	59197
Adult	55978
Senior	55763

5. Dashboard in Power BI

Finally, we built an interactive dashboard in **Power BI** to present insights visually.



6. Business Recommendations

- **Boost Subscriptions** – Promote exclusive benefits for subscribers.
- **Customer Loyalty Programs** – Reward repeat buyers to move them into the “Loyal” segment.
- **Review Discount Policy** – Balance sales boosts with margin control.
- **Product Positioning** – Highlight top-rated and best-selling products in campaigns.
- **Targeted Marketing** – Focus efforts on high-revenue age groups and express-shipping users.